

Measures of Reliance and Compliance in Aided Visual Scanning

Joachim Meyer, Tel Aviv University, Tel Aviv, Israel, and Rebecca Wiczorek and Torsten Günzler, Technische Universität Berlin, Berlin, Germany

Objective: We study the dependence or independence of reliance and compliance as two responses to alarms to understand the mechanisms behind these responses.

Background: Alarms, alerts, and other binary cues affect user behavior in complex ways. The suggestion has been made that there are two different responses to alerts—compliance (the tendency to perform an action cued by the alert) and reliance (the tendency to refrain from actions as long as no alert is issued). The study tests the degree to which these two responses are indeed independent.

Method: An experiment tested the effects of the positive and negative predictive values of the alerts (PPV and NPV) on measures of compliance and reliance based on cutoff settings, response times, and subjective confidence.

Results: For cutoff settings and response times, compliance was unaffected by the irrelevant NPV, whereas reliance depended on the irrelevant PPV. For subjective estimates, there were no significant effects of the irrelevant variables.

Conclusion: Results suggest that compliance is relatively stable and unaffected by irrelevant information (the NPV), whereas reliance is also affected by the PPV. The results support the notion that reliance and compliance are separate, but related, forms of trust.

Application: False alarm rates, which affect PPV, determine both the response to alerts (compliance) and the tendency to limit precautions when no alert is issued (reliance).

Keywords: trust, warnings, alerts, reliance, compliance, automation, signal detection theory, confidence

INTRODUCTION

Responses to binary alerts, alarms, and dynamic warnings have attracted considerable interest in recent years. One issue that arises in this context is the question of whether the trust in such systems is a single entity or whether there are actually two different forms of trust in binary cues (Meyer, 2004). One of them is *compliance*, which is the degree to which the binary cue, when it is present, causes the operator to act in accordance to the cue. Another type of trust, referred to as *reliance*, is the degree to which operators dare to avoid taking precautions when the binary indicator does not point to a signal.

A number of studies have shown that the two responses are differentially affected by different variables (Bahner, Elepfandt, & Manzey, 2008; Bitan & Meyer, 2007; Meyer, 2001; Yamada & Kuchar, 2006). For instance, McBride, Rogers, and Fisk (2011) showed that with greater workload participants tended to comply more with automation, but workload had no effect on reliance. Other studies, however, showed similar effects on both responses (e.g., Botzer, Meyer, & Parmet, 2013).

An analysis of the normative responses to cues (Meyer, 2004) reveals that compliance should be affected by the positive predictive value (PPV), whereas reliance should be affected by the negative predictive value (NPV). The PPV is the probability that there is a signal that needs to be detected (S), given that there was an alert (s), that is, $p(S|s)$. It depends on $p(s|S)$, which is the probability that a cue appears, given that there is a signal and is the hit rate ($pHit$) in signal detection theory (SDT). It also depends on $p(s|N)$, the probability of an alert (s), given that there is no signal (N), which is the false alarm rate (pFA) in SDT. These two probabilities are weighted with the base rate of a signal $p(S)$. From these probabilities one can compute PPV using Bayes's theorem:

Address correspondence to Joachim Meyer, Department of Industrial Engineering, Tel Aviv University, Ramat Aviv, Tel Aviv 69978, Israel; e-mail: jmeyer@tau.ac.il.

HUMAN FACTORS

Vol. 56, No. 5, August 2014, pp. 840–849

DOI: 10.1177/0018720813512865

Copyright © 2013, Human Factors and Ergonomics Society.

$$PPV = p(S|s) = \frac{p(s|S)p(S)}{p(s|S)p(S) + p(s|N)p(N)} \quad (1)$$

Similarly, one can compute the NPV, that is, how likely it is that there is no signal when no cue was given, $p(N|n)$. Whereas PPV decreases with more false alarms, NPV decreases with more missed detections by the cueing system.

Several studies have looked at the effects of the different types of errors on reliance and compliance. Rice (2009) had participants search aerial photographs for weapons. They were aided by a recommendation system that indicated whether or not it had detected a weapon. The aid was either false alarm prone or miss prone, with reliabilities between .55 and 1. Both automation errors affected reliance and compliance negatively. False alarms led to a relatively large reduction in compliance and a smaller, but significant, reduction in reliance, whereas misses led to a large reduction in reliance and a smaller, but significant, reduction in compliance. The author concluded that in addition to the primary selective effects, the two types of errors had some sort of nonselective crossover effects.

Other studies have revealed different results. Dixon and Wickens (2006) showed that false alarms in a system monitoring task affected both compliance and reliance. Missed detections, in contrast, affected only reliance. This finding was replicated in a later study (Dixon, Wickens, & McCarley, 2007). The authors showed that the effect of the two failures was not symmetrical, as PPV influenced both types of trust, whereas NPV did not. In another study Rice and McCarley (2011) found the same asymmetrical effect. The authors argued that false alarms are more salient than misses, and indeed the effect diminished when the system did not show false alarms as explicit messages.

The studies that revealed asymmetric effects on compliance and reliance all used extreme values for either PPV or NPV. Dixon et al. (2007) used, for example, $PPV = NPV = 1$ in the perfect automation condition, $PPV = .56$, $NPV = 1$ in the false alarm prone condition, and $PPV = 1$, $NPV = .56$ in the miss prone condition. The use of perfectly reliable indications in some conditions may have made the different reliabilities of

the two types of cues (alarms and nonalarms) particularly obvious and may have contributed to the asymmetric patterns found in the studies. It is not clear whether participants will equally respond to the two types of cues when both have only limited reliability, as is the case in most actual systems.

In our study we investigated whether such asymmetric effects exist when both types of system failures can occur and can influence trust simultaneously. We therefore varied PPV and NPV in a balanced manner without using extreme (PPV or $NPV = 1$) values. Conditions differed in one characteristic (the PPV or the NPV), whereas the other was kept constant. This allowed us to assess the independence of the two responses in a controlled setting—in this case a simulated inspection task, resembling the visual inspection of images for signals, as in airport luggage scanning. We used SDT to set the sensitivity and cutoff settings of the different cueing systems in this experiment.

One explanation for the asymmetry in the effects may be that false alarms are more easily detected than misses. Therefore operators will base their trust in automation more on the system characteristic that depends on false alarms (PPV) than on the system characteristic that depends on misses (NPV), causing PPV to have a stronger effect than NPV (Rice & McCarley, 2011). It is possible to test this hypothesis by keeping misses and false alarms equally salient and providing feedback after each trial.

Three possible patterns of results can emerge in our experiment:

1. The two responses may be independent (i.e., compliance will be unaffected by NPV, and reliance will be unaffected by PPV).
2. Both responses are related (i.e., compliance will also depend on NPV and reliance will also depend on PPV).
3. There is an asymmetry of the effects of NPV and PPV on responses, namely PPV (due to false alarms) will affect reliance, whereas NPV (due to misses) will have no effect on compliance.

A major difference between our study and previous studies on reliance and compliance is the way we measured the two types of trust. As

pointed out in a study on physicians' responses to clinical reminders (Vashitz et al., 2009), compliance and reliance should be computed relative to responses when no cues are available, rather than by comparing responses to different cues. In an experiment comparing alerting systems with different diagnostic values, the estimation of compliance and reliance relative to a control condition that has no alerting system allows us to distinguish between three different situations: First, if there is no effect of the diagnostic value of the alerting system, there may still be reliance or compliance if the conditions that had the alerting system differ significantly from the control condition. Second, if there is a significant difference between conditions with different diagnostic values of the alerting system, significant compliance or reliance may still exist in all conditions, but their strength may differ as a function of the validity of the alert. Third, there may only be compliance or reliance if the diagnostic value of the alerting system exceeds a certain value, and below it there may be no significant reliance or compliance. Therefore we compute the dependent variables as differences between the values in the unaided (control) blocks and the aided blocks.

METHOD

Participants

A total of 60 undergraduate students with self-reported normal or corrected-to-normal vision and without reported color vision deficiency participated in the study. Participants received 30 ILS (Israeli shekels, about \$8.40) for their participation and also took part in a lottery of 4 times 100 ILS (about \$28.00). Each point in the performance score was a lottery ticket for this lottery, so that participants had an incentive to collect as many points as possible.

Task Environment

Participants saw monochrome images, each showing a 3×3 matrix of single digits, displayed on 19-in. screens. In these images 39% of the pixels were inverted, resulting in a blurred image, as shown in Figure 1. Participants had to click a button labeled "target" when the target digit 3 was present in the image and had to click

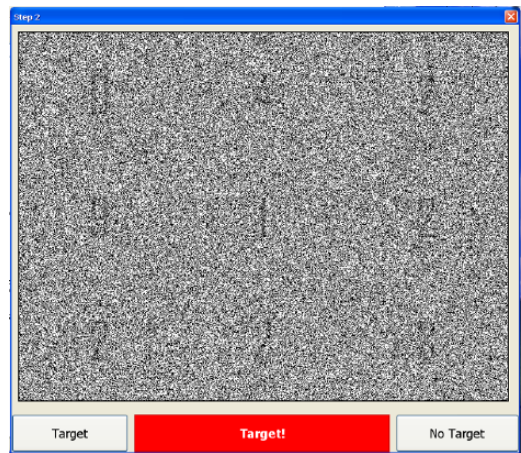


Figure 1. Example of a stimulus with a cue indicating the presence of a target (in this case the cue color is red).

a button labeled "no target" otherwise. Images appeared for 2 seconds and were not repeated. After the image disappeared, participants had 10 seconds to respond before they ran out of time. They received a reward for every correct answer and a penalty for every wrong answer. In blocks with cues participants saw either a red or a green rectangle labeled either "target" or "no target." The cues appeared for 4 seconds, 2 seconds prior to the image presentation, serving as a "target" or "no target" cue, respectively. The validity of the cues differed between the experimental conditions. After participants responded, they indicated how confident they felt about their decision on a 7-point Likert-type scale. After each trial they received feedback in the form of a short text message, informing them whether their decision was correct or not.

Procedure

Participants were randomly assigned to one of the four experimental conditions and one of six computer work stations. They read the instructions and then performed the four blocks of the experiment. The first and third block consisted of 40 trials each, and there were no cues (the unaided blocks). The results from these blocks served as control values to which the responses to cues could be compared. The second and the fourth blocks consisted of 100

TABLE 1: SDT Parameters of the Four Different Cueing Systems (c_{Cue} is the criterion and d'_{Cue} is the sensitivity of the cue)

	PPV	NPV	c_{Cue}	d'_{Cue}
NPV .75	.9	.75	0.44	1.93
NPV .95	.9	.95	−0.23	3.03
PPV .75	.75	.9	−0.44	1.93
PPV .95	.95	.9	0.23	3.03

Note. NPV = negative predictive value; PPV = positive predictive value.

trials each. In them binary cues aided performance, indicating the presence or absence of a target (the aided blocks). In all blocks, 50% of the trials were “signals” (i.e., in 50% of the trials the target digit 3 appeared on the screen). We debriefed participants at the end of the experiment and paid them.

Experimental Design

In the study we had NPV/PPV combinations (see Table 1) with each participant being in only one of these combinations. With the four PPV/ NPV conditions we investigated the influence of PPV (.95 or .75, with a constant NPV of .9) on reliance and the influence of NPV (.95 or .75 with a constant PPV of .9) on compliance. The experiment consisted of four blocks. Blocks 1 and 3 served as control blocks in which no alerts were available, and in Blocks 2 and 4 the participants received alerts according to the NPV/ PPV condition to which they were assigned. We compared performance in Blocks 2 and 4 to determine whether experience with the task affects compliance or reliance.

Dependent Variables

As a measure of participants’ performance, we used participants’ sensitivity ($d'_{equivalent}$ or d'_{eq}) in the control blocks (i.e., unaided blocks) and the combined sensitivity of the user and the information from the cue in the aided blocks. Sensitivity was computed as shown in Equation 2.

$$d'_{eq} = Z_{FA} - Z_{Hit} \tag{2}$$

We computed the sensitivity for all four blocks to determine if performance improved when cues of different quality were available.

We evaluated three different measures of compliance and reliance as a function of properties of the cueing system:

- 1. *Cutoff settings (C in terms of SDT)* (see Swets, Dawes, & Monahan, 2000). We measured participants’ responses to alerts by analyzing the differences in cutoff settings in unaided and aided blocks (for previous studies that used this measure, see Botzer et al., 2013; Maltz & Meyer, 2001; Meyer, 2001). A difference in the unaided and aided cutoff settings is evidence for compliance (when the cue indicates a signal) or reliance (when the alert indicates no signal). We computed participants’ cutoff settings C from the inverse normal distribution values (z) of $pHit$ and pFA . Lower cutoff settings indicate a more cautious cutoff (a greater tendency to identify a signal):

$$C = -0.5 (Z_{Hit} + Z_{FA}) \tag{3}$$

We computed the cutoff setting of the unaided block ($C_{Control}$) and one cutoff setting for red cues (C_{Alert}) and one cutoff setting for green cues ($C_{No-Alert}$) in the aided block. We computed the measure of *compliance* by subtracting the cutoff for red cues from the unaided cutoff:

$$Compliance = C_{Control} - C_{Alert} \tag{4}$$

Similarly, we computed the measure for *reliance* by subtracting the unaided cutoff from the cutoff for green cues:

$$Reliance = C_{No-Alert} - C_{Control} \tag{5}$$

Both measures have positive values if operators express either reliance or compliance, with

larger values when compliance or reliance are stronger.

2. *Response times.* Responses are faster when decisions are easier and require less processing (e.g., Botzer et al., 2013). We limited our comparison to the response times for the trials in which participants correctly identified a signal or its absence in the control blocks and responses when participants correctly identified a signal following a red cue or its absence following a green cue in the aided blocks. If responses are faster with cues (in the aided blocks) than without cues (in the unaided blocks), the difference in the response time is an expression of compliance or reliance. We computed compliance and reliance scores for the response times (RTs) as follows:

$$\text{Compliance} = \text{RT}_{\text{Control}} - \text{RT}_{\text{Alert}} \quad (6)$$

$$\text{Reliance} = \text{RT}_{\text{Control}} - \text{RT}_{\text{No-Alert}} \quad (7)$$

When the difference value was 0, there was no compliance or reliance, and positive difference values mean that there was compliance or reliance.

3. *Subjective confidence ratings.* Participants indicated their confidence in their choice after each decision on a 7-point Likert-type scale. Compliance or reliance is evident when confidence ratings are higher in the aided trials than ratings for unaided trials. For the analysis of confidence ratings (CRs), we computed the differences between the values in the unaided and aided blocks:

$$\text{Compliance} = \text{CR}_{\text{Alert}} - \text{CR}_{\text{Control}} \quad (8)$$

$$\text{Reliance} = \text{CR}_{\text{No-Alert}} - \text{CR}_{\text{Control}} \quad (9)$$

Here, too, a difference value of 0 means no compliance or reliance, and larger values point to greater compliance or reliance.

RESULTS

Sensitivity (d'_{eq})

We analyzed the d'_{eq} values for the four blocks to assess participants' task performance. The d'_{eq} values should be much larger in the conditions in which one predictive value was .95 and

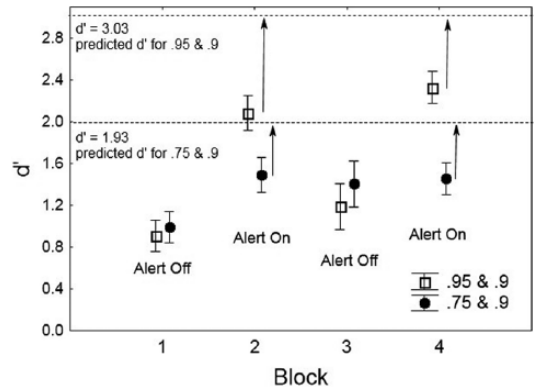


Figure 2. Predicted and observed sensitivity (d'_{eq}) values in the four blocks for the lower sensitivity system ($d'_{Cue} = 1.93$) and the higher sensitivity system ($d'_{Cue} = 3.03$). The arrows indicate the distance between the sensitivity achieved by the combined human-machine system and the sensitivity of the cueing system alone.

the other was .9, compared to the conditions in which one predictive value was .75 and the other was .9. As shown in Table 1, if participants respond only to the cues d'_{eq} will be 1.93 in the PPV .75 and NPV .75 conditions and 3.03 in the PPV .95 and NPV .95 conditions. Figure 2 shows the observed d'_{eq} values in the four conditions. We therefore analyzed these two conditions separately in two-way ANOVAs with the block (1 to 4) and the type of cue that was varied (PPV vs. NPV) as independent variables.

In the analyses of the two .75 conditions (PPV .75 and NPV .75) neither the effect of the type of cue nor the interaction Type of Cue \times Block were significant. There was, however, a significant main effect of the block, $F(3, 84) = 8.63$, $p = .0001$, $MSe = .187$, partial $\eta^2 = .24$, with d'_{eq} values of 0.99, 1.49, 1.41, and 1.46 for the four blocks. Since in Blocks 1 and 3 no cues were available, we should expect lower d'_{eq} values for them. The d'_{eq} in Block 1 was indeed lower compared to the other blocks, but there was no significant difference between Blocks 3 and 4. The increase of d'_{eq} from Block 1 to Block 3 might be due to a learning effect, as participants received no training before Block 1. Thus, when the cueing system has only limited validity in one of its indications, users with some

experience with the task did not significantly benefit from receiving the cues.

The two-way ANOVA for the .95 conditions also showed only a significant main effect of the block, $F(3, 84) = 72.09, p < .0001$, $MSe = .315$, partial $\eta^2 = .72$. Here the sensitivity in the two blocks in which cues were available was clearly higher than the sensitivity in the blocks in which no cues were available (with d'_{eq} values of 0.91, 2.08, 1.19, and 2.33 for the four blocks). Thus, performance in Blocks 2 and 4 (with cues) was clearly superior to the performance in Blocks 1 and 3.

In all four conditions, the d'_{eq} values for users who could rely on the combination of the image and cue did not reach the sensitivity level that could have been attained if participants would have responded only to the cues. Thus, one can argue that participants insufficiently trusted the cueing system, in the sense that more trust in this system could have helped them attain better discrimination performance.

Compliance and Reliance in Cutoff Settings

As mentioned earlier, the dependent variables in this and the following analyses were the differences between the control blocks (without cues) and the subsequent aided blocks (with cues). The factor block therefore had only two levels in this analysis (Blocks 2 and 4), where the value for Block 2 was the difference between Blocks 1 and 2 and the value for Block 4 was the difference between Blocks 3 and 4.

We conducted a separate analysis for compliance, comparing the two NPV conditions (NPV = .75 or .95) and a separate analysis for reliance, comparing the two PPV conditions (PPV = .75 or .95). We analyzed compliance and reliance with two-way ANOVAs with the predictive value (PPV/NPV = .75 or .95) as a between-subjects factor and block (Blocks 2 and 4) as a within-subject factor.

The upper-left panel of Figure 3 shows the results for the analysis of compliance. All values are significantly larger than 0 (as can be seen by comparing the standard error whiskers to the 0 value). The only significant effect in the analysis was the main effect of the block, $F(1, 28) = 4.94, p = .035$, $MSe = .313$, partial $\eta^2 = .15$. Compliance overall increased from Block 2 to Block 4.

The parallel analysis of reliance showed somewhat different results, as depicted in the upper-right panel of Figure 3. Here we found a significant main effect of the PPV value, $F(1, 28) = 8.50, p = .007$, $MSe = .671$, partial $\eta^2 = .233$. When PPV = .75, the reliance was not significantly different from 0. Thus in this case participants did not use the absence of a cue to determine that a situation was safe. In contrast, there was reliance when PPV = .95. No other effect was significant. Reliance developed very quickly when the system had a high PPV, as there was no effect of the block.

Compliance and Reliance in Response Times

We analyzed compliance and reliance in RTs with two separate two-way ANOVAs, parallel to the analyses we used for the cutoff settings. The middle panel of Figure 3 shows the results of the analyses. Only the main effect of the block was significant, $F(1, 28) = 5.00, p = .033$, $MSe = .171$, partial $\eta^2 = .152$. The compliance in Block 2 was significantly larger ($0.52, SE = 0.104$) than in Block 4 ($0.29, SE = 0.100$). The NPV value had no significant effect on the compliance as measured in RTs, nor was there an interaction with the block.

The only significant effect in the analysis of reliance was the main effect of the PPV value, $F(1, 28) = 5.27, p = .029$, $MSe = .792$, partial $\eta^2 = .159$. Although there was practically no reliance when PPV was .75 ($0.42, SE = 0.162$), it was high when PPV was .95 ($0.94, SE = 0.162$). Thus the irrelevant PPV affected participants' RTs when deciding that an item was intact.

Compliance and Reliance in Confidence Ratings

We analyzed compliance and reliance in subjective CRs with two separate two-way ANOVAs (see the bottom panels of Figure 3 for the results). For compliance there was only a main effect of the block, $F(1, 28) = 10.11, p = .004$, $MSe = .313$, partial $\eta^2 = .159$, with significantly more compliance in Block 2 ($0.46, SE = 0.127$) than in Block 4 ($0.002, SE = 0.132$), in which there was no significant compliance any more.

There were no significant effects in the parallel analysis of reliance. Overall, there was no reliance

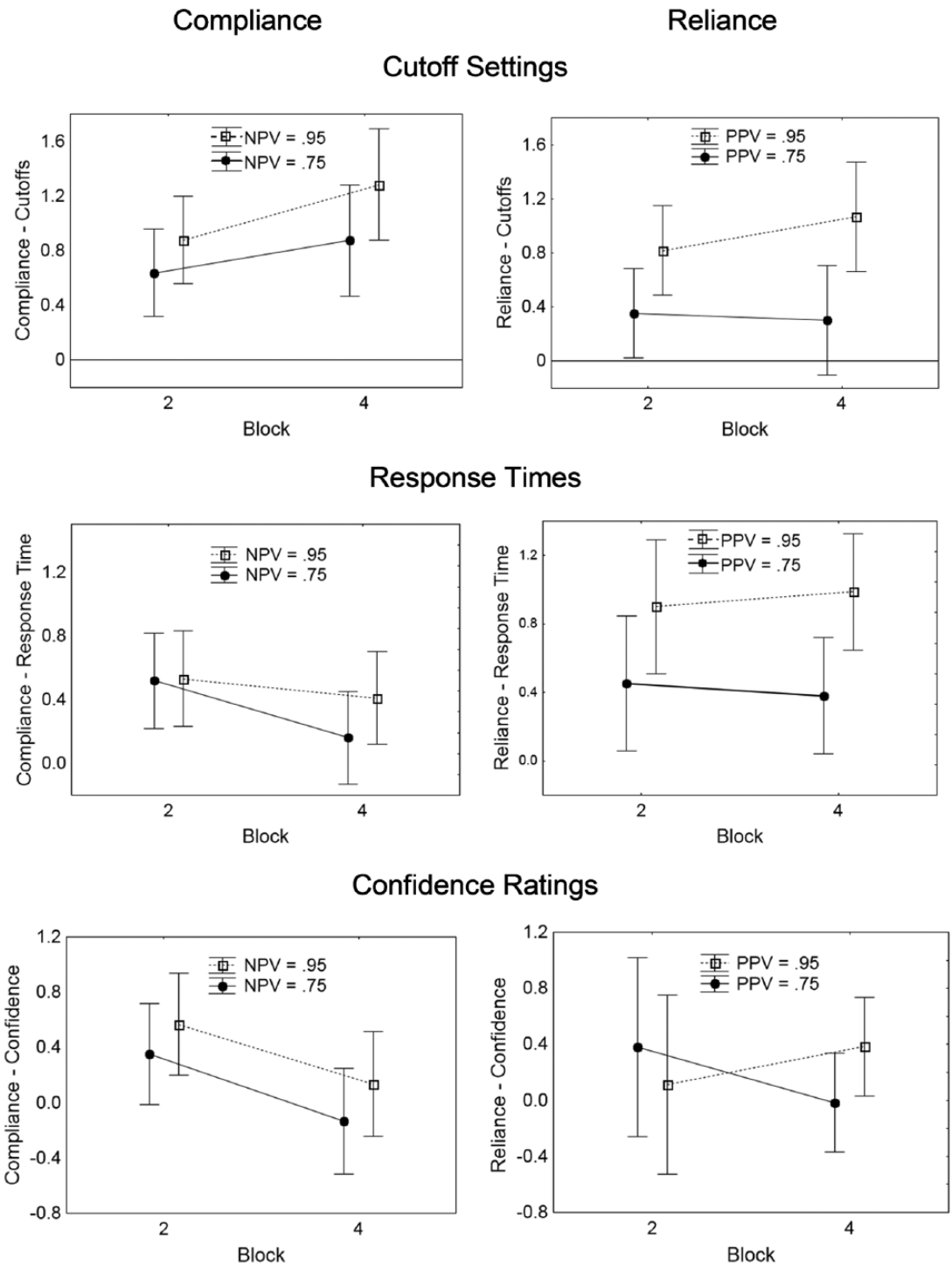


Figure 3. Compliance and reliance results as measured through cutoff settings, response times, and subjective confidence ratings. Vertical bars denote 95% confidence intervals.

in terms of confidence, and neither the PPV value nor the block had any effect on reliance.

DISCUSSION

In our experiment we assessed user responses to different cues in a visual scanning task. Our results show that the performance in this task does not reach the level of performance that could have been attained if the task would have been done automatically (i.e., if the alerting system would have made the categorization decisions). Involving the human operator in the task did not lead to better performance of the task, compared to relying entirely on the automation to perform the task. Participants gave insufficient weight to the information from the cue when the cue was only partly diagnostic (when NPV or PPV was .75) and when it was highly valid (NPV or PPV was .95). Some improvement in the performance in the unaided blocks occurred over time, but the responses to cues were relatively unaffected by learning.

We mainly intended to assess the relation between the two aspects of trust in a system (identified by Meyer, 2004)—the users' compliance with the system and the reliance on it. The two aspects could theoretically be two expressions of the same underlying trust. Alternatively, they could be two entirely independent types of trust. Finally, they can be somewhat related, with either affecting the other or with only one affecting the other. We tested compliance and reliance with three measures—the cutoff settings participants chose, the RTs, and the subjective confidence levels. The results for cutoff settings and RTs were similar. In both measures compliance increased significantly over time. No such effect existed for the measures of reliance. In the analyses of reliance, however, the main effect of the PPV value was significant for both the cutoff settings and the RTs.

The results for the CRs were different. Compliance significantly decreased over time, and there were no significant effects in the analysis of reliance. Considering these findings and studies on the numerous factors influencing such ratings (Bass, Baumgart, & Shepley, 2013; de Visser & Parasuraman, 2011; Morris, 1990; Paese & Sniezek, 1991), we conclude that subjective CRs

may not represent a suitable indicator of compliance and reliance.

Our results for the cutoff settings and RTs support the notion that reliance and compliance, as expressed in these measures, are separate but related types of trust. Reliance was affected by PPV. In fact, when PPV was .75, there was no evidence for reliance. Participants in this condition apparently did not rely on cues indicating the absence of a signal for deciding if a situation was intact. Reliance existed only when the PPV was high. Thus, for a user to trust a system when it indicates that everything is fine, the user seems to consider the likelihood that the system detects a failure if one exists.

In contrast, differences in NPV had no significant effects on compliance. Apparently the factor that affected the operators' response was the question of how much the system could be trusted when it issued an alert. Our relatively small sample prevents us from concluding that NPV has no influence on compliance at all. However, our study shows that PPV clearly has a stronger influence than NPV. As PPV decreases participants reduce both their compliance with alarms and their reliance on the system when no alarm is present.

It seems therefore that the distinction between reliance and compliance has some value, as they are not two expressions of the same underlying trust in the system. However, the two types of trust are not independent: A relatively high level of PPV is a precondition for reliance, but compliance does not require a relatively high level of NPV.

It is not clear what causes this asymmetry in the responses to information. Rice and McCarley (2011) suggested that the asymmetry may be due to the greater salience of false alarm events. Thus people are more aware of the failures of the system when it falsely issues an alert. They are less likely to notice instances when the system fails to detect a signal. This explanation is particularly plausible when only some of the events are accompanied by a cue (the alert is presented only when a signal is detected, otherwise it is silent).

In our study the different cues were visible in all trials in blocks with alerts, either as red or as green indicators. In addition, participants received

feedback after each trial, informing them about system failures. Thus, at least from the perspective of information presentation, the conditions with and without cues did not differ. Hence, the salience hypothesis, at least in its simpler form, does not account for our findings.

It might, however, be possible to maintain a salience hypothesis that is based on the apparent salience, rather than on the actual one. Participants may for some reason attend more closely to events for which an alert is issued than to events for which no alert is issued. Consequently, failures in the former type of events may be more vivid and affect responses stronger than failures when no alert is given.

An alternative explanation may be that people are more afraid of causing omission errors (to rely on the system and refrain from necessary action when the system indicates that no signal is present) than commission errors (conduct an unnecessary action when the system generates a false alarm). As a result the reliance component may be more vulnerable to any type of system failure, and reliance decreases whenever users' trust in the system diminishes, irrespective of if this is due to a low NPV or PPV. Gérard and Manzey (2010) found that the reduction of PPV from .9 to .7 decreased compliance about 25%. A similar decrease of about 25% in the reliance appeared when NPV was reduced from .98 to .93. Thus, reliance decreased disproportionately strong when the system provided only a few misses, whereas a larger decrease in PPV was necessary to reduce compliance to a comparable degree. These findings suggest that participants have a strong tendency to avoid omission errors if possible.

In summary, an asymmetry exists between reliance and compliance. Reliance seems to be a less robust phenomenon than compliance, and for it to appear, the alerts the system issues need to have a high probability of being justified. Thus the proportion of unjustified alerts affects both reliance and compliance, whereas the proportion of missed detections affects only reliance. This asymmetry is not limited to partly perfect systems (where either the PPV or the NPV is perfect), but it also exists in systems with both types of errors. It is evident in performance measures of trust, but not in subjective

CRs. These findings should be considered in theories of trust in alerts and in the development of predictive models of user responses to alerts in a given setting.

ACKNOWLEDGMENTS

This paper was funded with Israel Science Foundation (ISF) Grant 670/09. We thank Noa Dayan for her help in collecting the data and Shai Kerer for his programming services. Parts of this study were presented at the 56th annual meeting of the Human Factors and Ergonomics Society and were published as Wiczorek, Meyer, and Günzler (2012).

KEY POINTS

- We tested the relation between reliance and compliance responses to alerts in an aided binary categorization task.
- We measured response times, probabilities of responses, and subjective confidence ratings.
- False alarms affected both reliance and compliance when measured with RTs and probabilities, whereas missed detections affected only reliance.
- No evidence for reliance and compliance existed for subjective confidence ratings.

REFERENCES

- Bahner, J. E., Elepfandt, M., & Manzey, D. (2008). Misuse of diagnostic aids in process control: The effects of automation misses on complacency and automation bias. In *Proceedings of the 52nd Meeting of the Human Factors and Ergonomics Society* (pp.). Santa Monica, CA: Human Factors and Ergonomics Society.
- Bass, E. J., Baumgart, L. A., & Shepley, K. K. (2013). The effect of information analysis automation display content on human judgment performance in noisy environments. *Journal of Cognitive Engineering and Decision Making*, 7, 49–65.
- Bitan, Y., & Meyer, J. (2007). Self-initiated and respondent actions in a simulated control task. *Ergonomics*, 50, 763–788.
- Botzer, A., Meyer, J., & Parmet, Y. (2013). Mental effort in binary categorization aided by binary cues. *Journal of Experimental Psychology: Applied*, 19, 39–54.
- de Visser, E., & Parasuraman, R. (2011). Adaptive aiding of human-robot teaming: Effects of imperfect automation on performance, trust, and workload. *Journal of Cognitive Engineering and Decision Making*, 5, 209–231.
- Dixon, S. R., & Wickens, C. D. (2006). Automation reliability in unmanned aerial vehicle control: A reliance-compliance model of automation dependence in high workload. *Human Factors*, 48, 474–486.
- Dixon, S. R., Wickens, C. D., & McCarley, J. S. (2007). On the independence of compliance and reliance: Are automation false alarms worse than misses? *Human Factors*, 49, 564–572.
- Gérard, N., & Manzey, D. (2010). Are false alarms not as bad as supposed after all? A study investigating operators' responses

- to imperfect alarms. In D. de Waard, A. Axelsson, M. Berglund, B. Peters, & C. Weikert (Eds.), *Human factors: A system view of human, technology and organisation* (pp. 55–70). Maastricht, Netherlands: Shaker.
- Maltz, M., & Meyer, J. (2001). Use of warnings in an attentionally demanding detection task. *Human Factors*, 43, 217–226.
- McBride, S. E., Rogers, W. A., & Fisk, A. D. (2011). Understanding the effect of workload on automation use for younger and older adults. *Human Factors*, 53, 672–686.
- Meyer, J. (2001). Effects of warning validity and proximity on responses to warnings. *Human Factors*, 43, 563–572.
- Meyer, J. (2004). Conceptual issues in the study of dynamic hazard warnings. *Human Factors*, 46, 196–204.
- Morris, C. C. (1990). Retrieval processes underlying confidence in comprehension judgements. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16, 223–232.
- Paese, P. W., & Snizek, J. A. (1991). Influences on the appropriateness of confidence in judgment: Practice, effort, information, and decision-making. *Organizational Behavior and Human Decision Processes*, 48, 100–130.
- Rice, S. (2009). Examining single- and multiple process theories of trust in automation. *Journal of General Psychology*, 136, 303–319.
- Rice, S., & McCarley, J. S. (2011). Effects of response bias and judgment framing on operator use of an automated aid in a target detection task. *Journal of Experimental Psychology: Applied*, 17, 320–331.
- Swets, J. A., Dawes, R. M., & Monahan, J. (2000). Psychological science can improve diagnostic decisions. *Psychological Science in the Public Interest*, 1, 1–26.
- Vashitz, G., Meyer, J., Parmet, Y., Peleg, R., Goldfarb, D., Porath, A., & Gilutz, H. (2009). Defining and measuring physician's responses to clinical reminders. *Journal of Biomedical Informatics*, 42, 317–326.
- Wiczorek, R., Meyer, J., & Günzler, T. (2012). On the relation between reliance and compliance in an aided visual scanning task. In *Proceedings of the Human Factors and Ergonomics Society 56th Annual Meeting* (pp. 253–257). Santa Monica, CA: Human Factors and Ergonomics Society.
- Yamada, K., & Kuchar, J. K. (2006). Preliminary study of behavioral and safety effects of driver dependence on a warning system in a driving simulator. *IEEE Transactions on Systems, Man, and Cybernetics, Part A: Systems and Humans*, 36, 602–610.
- Joachim Meyer is professor at the Department of Industrial Engineering at Tel Aviv University. He earned his PhD in industrial engineering and management from Ben-Gurion University of the Negev, Israel.
- Rebecca Wiczorek is a postdoctoral research fellow at the Institute of Psychology and Ergonomics, Technische Universität Berlin. She earned her PhD in human factors from the Technische Universität Berlin in 2012.
- Torsten Günzler is a PhD student at the Center for Human-Machine Studies at the Technische Universität Berlin, Germany.

Date received: March 27, 2013

Date accepted: October 14, 2013